

(12) **EUROPEAN PATENT SPECIFICATION**

(45) Date of publication and mention
of the grant of the patent:
05.11.1997 Bulletin 1997/45

(51) Int Cl.⁶: **H04L 29/06, H04L 12/28**

(86) International application number:
PCT/US94/01177

(21) Application number: **94914690.6**

(87) International publication number:
WO 94/18776 (18.08.1994 Gazette 1994/19)

(22) Date of filing: **02.02.1994**

(54) **MULTIMEDIA DISTRIBUTION SYSTEM**
MULTIMEDIAVERTEILUNGSSYSTEM
SYSTEME DE DISTRIBUTION MULTIMEDIA

(84) Designated Contracting States:
AT BE CH DE DK ES FR GB GR IE IT LI LU MC NL
PT SE

(74) Representative: **Hanna, Peter William Derek et al**
Tomkins & Co.,
5 Dartmouth Road
Dublin 6 (IE)

(30) Priority: **03.02.1993 US 13009**
08.12.1993 US 164407

(56) References cited:
• **IEEE NETWORK: THE MAGAZINE OF**
COMPUTER COMMUNICATIONS., vol.4, no.6,
November 1990, NEW YORK US pages 32 - 49,
XP172741 T.D.C.LITTLE ET AL 'NETWORK
CONSIDERATIONS FOR DISTRIBUTED
MULTIMEDIA OBJECT COMPOSITION AND
COMMUNICATION'
• **IEEE JOURNAL ON SELECTED AREAS IN**
COMMUNICATION., vol.9, no.9, December 1991,
NEW YORK US pages 1368 - 1382, XP267529
T.D.C.LITTLE ET AL 'MULTIMEDIA
SYNCHRONIZATION PROTOCOLS FOR
BROADBAND INTEGRATED SERVICES'

(43) Date of publication of application:
29.11.1995 Bulletin 1995/48

(73) Proprietor: **NOVELL, INC.**
Orem, UT 84057-2399 (US)

(72) Inventors:
• **NELSON, David, L.**
Framingham, MA 01701 (US)
• **UPPALURU, Premkumar**
North Andover, MA 01845 (US)
• **ROMANO, Pasquale**
Boston, MA 02113 (US)
• **KLEIMAN, Jeffrey, L.**
Lexington, MA 02173 (US)

Note: Within nine months from the publication of the mention of the grant of the European patent, any person may give notice to the European Patent Office of opposition to the European patent granted. Notice of opposition shall be filed in a written reasoned statement. It shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European Patent Convention).

EP 0 683 951 B1

Description

Background of the Invention

This invention relates to the management of digitized media stream data, e.g., digitized video, and particularly relates to the capture, storage, distribution, access and presentation of digital video within a network computing environment.

Extensive technological advances in microelectronics and digital computing systems have enabled digitization of a wide range of types of information; for example, digital representations of text, graphics, still images and audio are now in widespread use. Advances in compression, storage, transmission, processing and display technologies have recently provided the capabilities required to extend the field of digitization to additionally include video information.

Conventionally, digitized audio and video are presented on, for example, a computer system or network by capturing and storing the audio and video streams in an interleaved fashion, i.e., segments of the two streams are interleaved. This requires storage of the digital audio and video in a single stream storage container, and further requires retrieving chunks of interleaved audio and video data at an aggregate rate which matches the nominal rate of an active presentation sequence. In this way, one unit of video (say, a frame) is physically associated in storage with one unit of audio (say, a corresponding 33 msec clip), and the two are retrieved from storage as a unit. Sequences of such audio and video units are then provided to a presentation and decoder digital subsystem in an alternating fashion, whereby each audio and video unit of a pair is provided in sequence.

Computer systems that provide this audio and video management functionality typically include digital compression/decompression and capture/presentation hardware and software, and digital management system software, all of which is based upon and depends upon the interleaved format of the audio and video streams it processes.

Currently, handling of audio and video in a network environment is also based on a scheme in which capture, storage, and transmission of audio and video must be carried out using interleaved audio and video streams. This interleaving extends to the transmission of audio and video streams across the network in an interleaved format within transmission packets.

Synchronization of audio with video during an active presentation sequence is conventionally achieved by initially interleaving the audio and video streams in storage and then presenting audio and video chunks at the nominal rate specified for an active presentation sequence.

In "Time Capsules: An Abstraction for Access to continuous-Media Data," by Herrtwich, there is disclosed a frame-work based on time capsules to describe how timed data shall be stored, exchanged, and ac-

cessed in real-time systems. When data is stored into such a time capsule, a time stamp and a duration value are associated with the data item. The time capsule abstraction includes the notion of a clock for ensuring periodic data access that is typical for continuous-media applications. By modifying the parameters of a clock, presentation effects such as time lapses or slow motion may be achieved.

While the Herrtwich disclosure provides a time capsule abstraction for managing time-based data, the disclosure does not provide any technique for synchronizing time-based data based on the time capsule abstraction, and does not address the requirements of time-based data management in a network environment. Furthermore, the disclosure does not address processing of time-based data streams as a function of their interleaved format or manipulation of that format.

Further disclosures defining the general state of the art are to be found in IEEE Network: The Magazine of Computer Communications, vol. 4, No. 6, November 1990 - "Network considerations for distributed multimedia object composition and communication" by T.D.C. Little et al, and IEEE Journal on Selected Areas in Communication, vol. 9, No. 9, December 1991 - "Multimedia synchronization protocols for broadband integrated services" by T.D.C. Little et al.

Summary of the Invention

The invention provides a computer-based media data processor in accordance with claims 15, and 12 which follow.

In general, in one aspect, the invention features a computer-based media data processor for controlling the computer presentation of digitized continuous time-based media data composed of a sequence of presentation units. Each presentation unit is characterized by a prespecified presentation duration and presentation time during a computer presentation of the media data and is further characterized as a distinct media data type. In the processor of the invention, a media data input manager retrieves media data from a computer storage location in response to a request for computer presentation of specified presentation unit sequences, and determines the media data type of each presentation unit in the retrieved media data. The input manager then designates each retrieved presentation unit to a specified media data presentation unit sequence based on the media data type determination for that presentation unit. The input manager then assembles a sequence of presentation descriptors for each of the specified presentation unit sequences, each descriptor comprising media data for one designated presentation unit in that sequence, and each sequence of presentation descriptors being of a common media data type; and then associates each presentation descriptor with a corresponding presentation duration and presentation time, based on the retrieved media data. Finally, the input

manager links the presentation descriptors of each sequence to establish a progression of presentation units in that sequence.

A media data interpreter of the invention indicates a start time of presentation processing of the presentation descriptor sequences, and accordingly, maintains a current presentation time as the sequences are processed for presentation. The interpreter counts each presentation unit in the media data sequences after that unit is processed for presentation, to maintain a distinct current presentation unit count for each sequence, and compares for each of the presentation unit sequences a product of the presentation unit duration and the current presentation unit count of that sequence with the current presentation time after each presentation unit from that sequence is processed for presentation. Based on the comparison, the interpreter releases a presentation unit next in that presentation unit sequence to be processed for presentation when the product matches the current presentation time count, and deletes a presentation unit next in that presentation unit sequence when the product exceeds the current presentation time count.

In general, in another aspect, the invention features a media data processor for controlling transmission of digitized media data in a packet switching network. Such a network comprises a plurality of client computer processing nodes interconnected via packet-based data distribution channels. In the invention, a remote media data controller receives from a client processing node a request for presentation of specified presentation unit sequences, and in response to the request, retrieves media data from a corresponding media access location. A remote media data input manager of the invention then determines the media data type of each presentation unit in the retrieved media data, and designates each retrieved presentation unit to a specified media data presentation unit sequence based on the media data type determination for that presentation unit. Then the input manager assembles a sequence of presentation descriptors for each of the specified presentation unit sequences, each descriptor comprising media data for one designated presentation unit in that sequence, and all presentation descriptors in an assembled sequence being of a common media data type. The interpreter associates each presentation descriptor with a corresponding presentation duration and presentation time, based on the retrieved media data; and finally, links the descriptors in each assembled sequence to establish a progression of presentation units in each of the specified presentation unit sequences.

A remote network media data manager of the invention assembles transmission presentation unit packets each composed of at least a portion of a presentation descriptor and its media data, all presentation descriptors and media data in an assembled packet being of a common media data type; and releases the assembled packets for transmission via the network to the client

processing node requesting presentation of the specified presentation unit sequences.

A local media data controller of the invention transmits the presentation unit sequence request to the remote media data controller from the client processing node, and controls starting and stopping of sequence presentation in response to user specifications.

A local network media data manager of the invention receives at the client processing node the transmission presentation unit packets via the network, and designates a presentation unit sequence for each presentation descriptor and its media data in the received packets to thereby assemble the presentation descriptor sequences each corresponding to one specified presentation unit sequence, all presentation descriptors in an assembled sequence being of a common media data type. Then the local network media data manager links the descriptors in each assembled sequence to establish a progression of presentation units for each of the presentation unit sequences.

A local media data interpreter of the invention accepts the assembled presentation descriptor sequences one descriptor at a time and releases the sequences for presentation one presentation unit at a time. In this process, the local interpreter indicates a start time of presentation processing of the sequences, and accordingly, maintains a current presentation time as the descriptor sequences are processed for presentation. Based on the presentation duration of each presentation unit, the interpreter synchronizes presentation of the specified presentation unit sequences with the current presentation time.

In preferred embodiments, the specified media data presentation unit sequences comprise a video frame sequence including a plurality of intracoded video frames; preferably, each frame of the video frame sequence comprises an intracoded video frame, and more preferably, the video frame sequence comprises a motion JPEG video sequence and an audio sequence. In other preferred embodiments, each of the plurality of intracoded video frames comprises a key frame and is followed by a plurality of corresponding non-key frames, each key frame including media data information required for presentation of the following corresponding non-key frames.

In other preferred embodiments, synchronization of presentation of the specified presentation unit sequences is accomplished by the local media data interpreter by comparing for each of the presentation descriptors in each of the presentation descriptor sequences the presentation time corresponding to that descriptor with the currently maintained presentation time. Based on this comparison, the interpreter releases a next sequential presentation unit to be processed for presentation when the corresponding presentation time of that descriptor matches the current presentation time, and deletes a next sequential presentation unit to be processed for presentation when the current presentation time ex-

ceeds the corresponding presentation time of that descriptor.

In other preferred embodiments, synchronization of presentation of the specified presentation unit sequences is accomplished by the local media data interpreter by counting each presentation descriptor in the sequences after that presentation unit is released to be processed for presentation, to maintain a distinct current presentation unit count for each sequence. Then, the interpreter compares for each of the presentation unit sequences a product of the presentation unit duration and the current presentation descriptor count of that sequence with the currently maintained presentation time after a presentation unit from that sequence is released to be processed for presentation. Based on the comparison, the interpreter releases a next sequential presentation unit in that presentation unit sequence when the product matches the currently maintained presentation time, and deletes a next sequential presentation unit in that presentation unit sequence when the product exceeds the currently maintained presentation time.

In other preferred embodiments, the remote media data controller of the invention receives from the local media data controller, via the network, an indication of a specified presentation data rate at which the specified presentation unit sequences are to be transmitted via the network to the client node. The media data retrieved comprises a plurality of storage presentation unit sequences stored in a computer storage location, each storage presentation unit sequence composed of presentation units corresponding to a specified presentation unit sequence and all presentation units in a storage presentation unit sequence being of a common media data type. The remote media data input manager designates each of a portion of the presentation unit descriptors as the descriptor sequences are assembled, the portion including a number of descriptors based on the specified presentation data rate, each designated descriptor comprising null media data, to thereby compose the presentation descriptor sequences with only a portion of storage presentation unit media data. With this designation, the specified presentation unit sequences attain the specified presentation data rate of transmission.

In the invention, the separation of media streams and distinctly formatting of network transmission packets for each stream provides an opportunity and the facility to examine, process, and make transmission decisions about each stream and each presentation unit independent of other streams and presentation units. As a result, the media processor of the invention can make presentation decisions about a given presentation unit independent of the other units in the corresponding stream, and can make those decisions "on-the-fly". This capability provides for real time scaling and network load adjustment as a stream is retrieved, processed, and transmitted across the network.

Further aspects, features, and advantages of the in-

vention are set forth in the following specification and the claims.

Brief Description of the Drawing

Fig. 1 is a schematic diagram of media stream access and delivery points with which the digital video management system of the invention may interface; Fig. 2 is a schematic diagram of a stand-alone implementation of the digital video management system of the invention; Fig. 3 is a schematic diagram of a network implementation of the digital video management system of the invention; Fig. 4 is a schematic diagram of the local digital video management system manager modules of the invention; Fig. 5 is a schematic diagram illustrating the flow of media stream data between the stream I/O manager and stream interpreter modules of the local digital video management system manager of Fig. 4; Fig. 6 is a schematic flow chart illustrating presentation and capture scenarios carried out by the local digital video management system manager of Fig. 4; Fig. 7 is a schematic illustration of the translation from media stream storage format to token format carried out by the local digital video management system manager of Fig. 4; Fig. 8 is a schematic flow chart illustrating presentation and capture scenarios carried out by a digital video system used in conjunction with the local digital video management system manager scenarios of Fig. 6; Fig. 9 is a schematic diagram of the local digital video management system manager and the remote digital video management manager modules of the invention in a network implementation; Fig. 10 is a schematic diagram illustrating the flow of media stream data between the remote and local digital video management manager modules of the invention in a network implementation; Fig. 11A is a schematic flow chart illustrating presentation and capture scenarios carried out by the remote digital video management system manager of Fig. 9; Fig. 11B is a schematic flow chart illustrating presentation and capture scenarios carried out by the local digital video management system manager of Fig. 9; Fig. 12 is a schematic illustration of the translation of stream tokens of Fig. 7 into packet format.

Description of a Preferred Embodiment

Referring to Fig. 1, there is illustrated the digital video management system (DVMS) 10 of the invention. The DVMS provides the ability to capture, store, trans-

mit, access, process and present live or stored media stream data, independent of its capture or storage location, in either a stand-alone or a network environment. The DVMS accommodates media stream data, i.e., continuous, high data-rate, real-time data, including video, audio, animation, photographic stills, and other types of continuous, time-based media data. Throughout this description, the DVMS of the invention will be explained with reference to audio and video streams, but it must be remembered that any time-based media data stream may be managed in the system. In the DVMS, as shown in Fig. 1, media data may be accessed from, e.g., live analog capture, analog or digital file storage, or live digital capture from e.g., a PBX (private branch exchange) server, among other access points. The accessed media is managed by the DVMS for delivery to, e.g., a presentation monitor, a computer system for editing and presentation on the computer, a VCR/tape-printer, or digital storage, or sent to a PBX server.

Of great advantage, the DVMS management scheme is independent of any particular storage or compression technology used to digitize the data streams, and further, is independent of any particular communication protocols or delivery platform of a network in which the DVMS is implemented. Additionally, the DVMS is industry standards-based yet is flexible and standards-extensible, via its layered architecture, which incorporates multiple management platforms. Each of these features and advantages will be explained in detail in the discussion to follow.

Digital Video Management System Components

The DVMS of the invention is based on a technique whereby media data streams are handled and managed as distinct and separate media data streams in which there is no interleaving of media data. Here the term "stream" is meant to represent a dynamic data type, like video, as explained above, and thus, a stream consists of dynamic information that is to be produced and consumed in a computer system or network with temporal predictability. A stream contains a succession of sequences. Sequences can themselves contain sequences; in turn, each sequence contains a succession of segments. Streams, sequences and segments, as information identifiers, have no media type-specific semantics. Rather, they are convenient abstractions for specifying and organizing dynamic data types to be managed by the management system of the invention. An easily understood analogy to streams, sequences and segments is that of documents containing chapters, sections and sentences.

Streams are characterized by their media data type, e.g., audio, video, or animation data types. Sequences represent information that is meaningful to the user. For example, a video sequence may represent a video clip containing a video scene. Segments can be convenient "chunks" of data for editing and mixing that data. Seg-

ments may also represent units of data that are temporally linked, as when using a video compression scheme that produces key video frames and corresponding following difference video frames.

In the DVMS of the invention, streams that are intended for synchronous presentation can be grouped into a stream group of distinct constituent streams (i.e., without interleaving). Although constituent streams in such a stream group may be stored in an interleaved form within a storage container, the DVMS can dynamically coordinate separately stored streams; in either case, the system processes the streams distinctly, rather than in an interleaved fashion.

Segments of streams contain presentation units. A presentation unit is a unit of continuous, temporally-based data to be presented, and accordingly, has an associated presentation time and presentation duration. A presentation time indicates the appropriate point in the sequence of a presentation at which the associated presentation unit is to be played, relative to a time base for the ongoing presentation. A presentation duration indicates the appropriate interval of time over which the associated presentation unit is to be played in the ongoing presentation. Thus, a video presentation unit comprises a video frame, and an audio presentation unit comprises a number of sound samples associated with a frame duration.

As mentioned above, the DVMS may be implemented in a stand-alone computer system or a computer-based, packet switched network. Referring to Fig. 2, in a stand-alone computer system implementation 12, live or stored media streams are accessed and captured for presentation and editing on the stand-alone computer 14. The captured, and optionally edited media streams may then be delivered to a presentation monitor or to a VCR tape printer utility.

Referring to Fig. 3, a packet switching network in which the DVMS is implemented comprises desktop computer systems 18 which are linked via a packet switching network 80, which is controlled by the DVMS network implementation 16. The network 80 may comprise a local area network (LAN) or a wide area network (WAN), or a combination of one or more LANs and WANs. The DVMS provides access to and capture of media streams from live analog video capture, e.g., a VCR or camcorder, a network, storage or PBX server, or one of the desktop computers, and in turn manages the transmission of the media stream data across the network back to any of the access points.

The digital video management system consists of a local DVMS manager and a remote DVMS manager. The local DVMS manager provides a client operating environment, and thus resides on a stand-alone computer or each client computer in a network, "client" here being defined as a computer system or one of the access points in a network that request media data; the remote DVMS manager provides a network operating environment, and thus resides on a network server. Th

local DVMS manager may be implemented on, for example, IBM-compatible personal computers running Microsoft® Windows™, to thereby provide high-level, industry-standard access to underlying digital video services. This local DVMS manager implementation may support, for example, the industry-standard Microsoft® digital video MCI API for application development. The local DVMS manager incorporates an efficient data-flow subsystem, described below, that is highly portable to other operating systems.

The DVMS system of the invention is preferably implemented as an application programming interface suite that includes interfaces for a computer programming application to include media data stream management capability within the application. Thus, the DVMS interfaces with an underlying programming application via interface calls that initiate media data stream functions within the realm of the programming application. Such an interface implementation will be understandable to those skilled in the art of C programming.

The remote DVMS manager acts to dynamically link a client and a server in the packet network environment. The architecture of this manager has the important advantage of supporting the ability to scale distinct, non-interleaved media data streams, as discussed in depth below. This ability to scale packet-based video, thereby creating scalable packet video, is a facility which permits adaptive bandwidth management for dynamic media data types in both LANs and WANs. The remote DVMS manager may be implemented as a Netware® Loadable Module, on, for example, the Novell Netware® operating system.

Local DVMS Manager

The local DVMS manager manages the access and capture of media data streams transparently, i.e., without impacting the functionality of the application program which requested that access and capture. The local DVMS manager works with a digital video system, implemented either in special purpose digital video hardware or in special purpose software-based emulation of the digital hardware.

Referring to Fig. 4, the local DVMS manager 20 consists of three modules: the stream controller 24, stream input/output (I/O) manager 26, and the stream interpreter 28. This modularity is exploited in the DVMS design to separate the flow of data in a media data stream from the flow of control information for that media stream through the system. Based on this data and control separation, stream data and stream control information are each treated as producing distinct interactions among the three manager modules, which operate as independent agents. The I/O manager, interpreter and controller agents are each mapped via the local DVMS manager to independently scheduable operating system processes with independent program control flow and data space allocation. The flow of media stream da-

ta is managed by the stream I/O manager 26 and the stream interpreter 28, while the flow of control information is managed by the stream controller 24. Each of these management functions is explained in detail below.

The stream I/O manager module 26 is responsible for the dynamic supply of media data streams, e.g., audio and video streams, from or to the stream interpreter. This module also provides efficient file format handling functions for the media data, if it is accessed via a storage file, e.g., a DVI® AVSS file. In a stand-alone implementation of the DVMS of the invention, the stream I/O manager provides retrieval and storage of media data streams from or to points of media access, such as digital or analog storage containers, while in a network implementation of the DVMS, as described below, the remote DVMS manager modules provide retrieval and storage at points of media access via the network. Most importantly, the stream I/O manager performs a translation from the representation of audio and video information as that information is stored to the corresponding dynamic computer-based representation. This translation is explained in detail below.

The stream interpreter module 28 is responsible for managing the dynamic computer-based representation of audio and video as that representation is manipulated in a stand-alone computer or a computer linked into a packet network. This dynamic management includes synchronization of retrieved audio and video streams, and control of the rate at which the audio and video information is presented during a presentation sequence. In addition, the stream interpreter module manages the capture, compression, decompression and playback of audio and video information. This module is, however, compression technology-independent and additionally is device-independent. Base services of a digital video subsystem, including, for example, hardware for capture and presentation functions, are preferably implemented to be accessed through a standard API suite of digital video primitives, which encapsulate any functions unique to a particular compression or device technology.

The following suite of primitive functions provide device-independent access to the base services of a digital video subsystem:

Open: Open a specified device, initialize it, and return a handle for further requests;

Close: Close a specified device and free up any associated resources;

Get_Capabilities: Query a device's capabilities, e.g., display resolutions, compression format, etc.;

Start: Start decoding and displaying data from a stream buffer;

Stop: Stop decoding and displaying data from a stream buffer;

Get_Info: Get information about the current status of a device;

Set_Info: Set information in the device attributes.

The stream controller module 24 is responsible for the control of video and audio capture and playback functions during user-directed applications. This control includes maintaining the dynamic status of video and audio during capture or playback, and additionally, providing presentation control functions such as play, pause, step and reverse. This module is accordingly responsible for notifying an active application of stream events during audio and video capture or playback. An event is here defined as the current presentation unit number, for which an indication would be made, or the occurrence of the matching of a prespecified presentation unit number with a current presentation unit number.

During the active playback of audio and video, or other dynamic media data streams, the stream I/O manager and the stream interpreter act as the time-based producer and consumer, respectively, of the data streams being played back. Conversely, during recording of a dynamic data stream, the stream interpreter acts as the time-based stream producer and the stream I/O manager acts as the time-based stream consumer. During both playback and recording, the I/O manager and the interpreter operate autonomously and asynchronously, and all data in an active stream flows directly between them via a well-defined data channel protocol. The stream controller asynchronously sends control messages to affect the flow of data between the I/O manager and the interpreter, but the controller does not itself participate in the flow of data. As discussed below, all data flow operations are handled using a minimal number of buffer copies between, for example, a disk or network subsystem and the digital video capture and presentation hardware.

This system design is particularly advantageous in that it provides for complete transparency with respect to the domain of the I/O manager and the interpreter, thereby providing the ability to extend the system to a network client/server configuration, as explained below. Moreover, this basic three-agent unit may be concatenated or recursed to form more complex data and control functionality graphs.

In the architecture of the local DVMS manager, the activity of one of the asynchronous agents, each time it is scheduled to run while participating in a stream flow, is represented as a process cycle. The rate at which an asynchronous agent is periodically scheduled is represented as the process rate for that agent, and is measured as process cycles per second. A process period is defined as the time period between process cycles. In order to maintain continuous data flow of streams between the stream I/O manager and the stream interpreter, the limiting agent of the two must process a process period's worth of presentation units within a given process cycle. In cases in which such process rates are not achieved, the local DVMS manager can control the flow rate, as explained below. The process rate for the

stream interpreter is close to the nominal presentation rate of the stream, i.e., in every process cycle, a presentation unit is processed. The stream I/O manager services several presentation units in every process cycle and thus, its process rate may be much lower than the presentation rate.

The modularity of the stream control functions provided by the stream I/O manager, interpreter and controller make the local DVMS manager architecture of the DVMS highly portable to most modern computer operating systems which support preemptive multitasking and prioritized scheduling. This architecture also provides for selective off-loading of the stream I/O manager and interpreter modules to a dedicated coprocessor for efficient data management. Most importantly, the highly decentralized nature of the manager architecture allows it to be easily adapted to LAN and WAN systems, as discussed below.

Referring to Fig. 5, when a computer-implemented with the DVMS of the invention requests access to audio or video streams, the following stream flow occurs. The stream I/O manager 26 module retrieves the requested streams from a stream input 30; this stream input comprises a storage access point, e.g., a computer file or analog video source. The stream I/O manager then separates the retrieved streams according to the specified file format of each stream. If two streams, e.g., audio and video streams, which are accessed were interleaved in storage, the stream I/O manager dynamically separates the streams to then transform them to distinct internal representations, each comprising a descriptor which is defined based on their type (i.e. audio or video). Once separated, the audio and video stream data are handled both by the stream I/O manager and the stream interpreter as distinct constituent streams within a stream group. The stream I/O manager 26 then exchanges the stream data, comprising sequences of presentation units, with the stream interpreter 28 via a separate queue of presentation units called a stream pipe 32, for each constituent stream; an audio stream pipe 33 is thus created for the audio presentation units, and a video stream pipe 31 is created for the video presentation units. Each audio stream (of a group of audio streams) has its own pipe, and each video stream has its own pipe. During playback of streams, the stream I/O manager continually retrieves and produces presentation units from storage and the stream interpreter continuously consumes them, via the stream pipes, and delivers them to a digital media data subsystem for, e.g., presentation to a user.

When retrieving a plurality of streams from an input 30 in which the streams are separated (not interleaved), the stream I/O manager retrieves and queues the streams' data in a round robin fashion, but does not perform any stream separation function. The stream interpreter processes these streams in the same manner as it processes those which are originally, interleaved. Thus, the stream I/O manager advantageously shields

the remainder of the system from the nature of the static container 30, and further "hides" the format of the storage container, as well as the way that logically coordinated data streams are aggregated for storage. Additionally, the details of the stream interpreter implementation, such as its hardware configuration, are "hidden" from the I/O subsystem; in fact, the only means of communication between the two agents is via the well-defined stream pipe data conduits.

Referring also to Fig. 6, during a presentation scenario, the stream controller 24 first initializes 36 the stream I/O manager 26 and stream interpreter 28, by creating active modules of them to begin processing streams, and then defines and indicates 38 a stream group and the corresponding constituent stream names. The stream I/O manager 26 then retrieves 40 the named streams from corresponding storage containers 30 and separates the streams, if stored in an interleaved fashion. If they were not interleaved, the streams are retrieved in a round-robin fashion. Once the streams are retrieved, the stream I/O manager converts 42 the streams to an internal computer representation of stream tokens, described below. Via the stream group indication 30, each stream token is identified with a stream and a stream group by the indication provided to the stream I/O manager by the stream controller. The I/O manager then buffers 44 the streams separately, each in a distinct stream pipe 32 for consumption by the stream interpreter 28; the stream controller provides control 46 of the stream group as it is enqueued.

Referring also to Fig. 7, the I/O manager stream translation 42 from storage representation to stream token representation is as follows. Typically, audio and video data is stored in an interleaved fashion on a disk and so upon retrieval are in an interleaved disk buffer, as in the Intel® AVSS file format. The disk buffers 100 consist of a sequence of stream group frames 105, each frame containing a header 106, a video frame 108, and an audio frame 110. A separate index table (not shown) containing the starting addresses of these stream group frames is maintained at the end of a file containing these frames. This index table permits random access to specifically identified stream group frames.

The disk buffers are retrieved by the I/O manager from the disk in large chunks of data, the size of each retrieved chunk being optimized to the disk track size, e.g., 64 K bytes each. The I/O manager examines each retrieved stream group frame header and calculates the starting addresses of each audio and video frames within the stream group frame. It also retrieves the time stamp information from the corresponding frames. A linked list of descriptors, called tokens 112, is then generated for the audio and video frames; each token represents an audio or video presentation unit 114 and the time stamp 116 for that unit. These tokens are continuously linked into a list representing the stream pipe. Thus, in the process described above, the stream I/O manager retrieves interleaved data from a disk, sepa-

rates the data into distinct streams, and constructs an internal representation of separated streams based on separate stream pipes, one for each stream.

Once the streams are enqueued in the stream pipes, the stream interpreter 28, having been initialized 36 by the stream controller 24, accepts and dequeues 48 the constituent stream tokens of presentation units. The debuffered streams are then scaled 50 and synchronized 52, based on control via the stream controller, which maintains 54 the status of the stream group. The scaling process will be described in detail below. The synchronized streams are then delivered to the digital presentation subsystem hardware.

The decompression scheme is based on the particular compression format of video frames, e.g., the motion JPEG video format. This format is one of a preferred class of video formats, in which each frame is intracoded, i.e., coded independently, without specification of other frames.

Referring to Fig. 8, the digital video system 120 receives streams from the stream interpreter and first decodes and decompresses 122 the stream data, each stream being processed separately. The decoded and decompressed data streams are then stored 124 in corresponding frame buffers, e.g., video and audio frame buffers. At the appropriate time, the stored data is converted 126 from its digital representation to a corresponding analog representation, and is delivered to a playback monitor and audio speakers. The various operations of the digital hardware subsystem are controlled by the stream interpreter via digital video primitives, as explained and described previously.

In the reverse operation, i.e., capture and storage of digital video and audio streams being processed by a computer system, the stream interpreter 28 captures the audio and video streams from the digital hardware subsystem digitizes 128 the audio and video signals, stores 130 the digitized signals in a buffer, and before passing the digitized streams to the stream interpreter, compresses and encodes 132 the video and audio data.

Based on the stream group control provided by the local stream controller, the stream interpreter generates 62 time stamps for the captured streams and using the time stamps, creates 64 corresponding stream tokens of video and audio presentation units with embedded time stamps. The stream tokens are then enqueued 66 to stream pipes 32 for consumption by the stream I/O manager 26.

The piped streams are accepted and dequeued 72 by the stream I/O manager 26, and then scaled. If the streams are to be stored in interleaved form, they are then interleaved 76, in a process which reverses the functionality depicted in Fig. 7. The streams are not required, of course, to be stored in such an interleaved form. Once the streams are interleaved, if necessary, the streams are stored in a corresponding storage container 30.

Each of the functions of the stream controller, stream I/O manager, and stream interpreter described in these scenarios may be implemented in hardware or software, using standard design techniques, as will be recognized by those skilled in the art.

Synchronization of Audio with Video

As mentioned in the presentation process described above, the digital video management system of the invention provides synchronization of audio to video, and in general, synchronization between any two or more dynamic streams being presented. This synchronization function is inherently required for the coordinated presentation of multiple real-time, continuous, high data-rate streams in a stream group. For example, the real-time nature of audio and video is derived from the presentation attributes of these dynamic data types, which have quite different presentation attributes; full motion video needs to be presented as 30 frames per second and high quality audio needs to be presented at 32,000 samples per second.

Furthermore, digital video and audio data streams have real-time constraints with respect to their presentation. The streams are usually continuous and last from 30 seconds-long (clips) to 2 hours-long (movies). Additionally, the streams typically consume from about 1 Mbit/sec to 4 Mbit/sec of storage capacity and transmission bandwidth, depending on the particular compression technology used for digitizing the stream. Thus, synchronization of differing data streams must accommodate the diverse temporal aspects of the streams to be synchronized.

The synchronization capability of the digital video management system of the invention is based on self-timing, and accordingly, self-synchronization, of data streams to be synchronized. This technique accommodates independent handling of multiple data streams which are together constituent streams of a stream group, even if the stored representations of the constituent streams are interleaved; the stream I/O manager separates interleaved streams before the stream interpreter synchronizes the streams. Alternatively, independent constituent streams may, however, be stored in separate file containers and be synchronized, before presentation, with a common reference time base.

Self-synchronization also provides the ability to prioritize one constituent stream over other streams in a stream group. For example, an audio stream may be prioritized over a video stream, thereby providing for scalable video storage, distribution and presentation rates, as discussed below. This feature is particularly advantageous because human perception of audio is much more sensitive than that of video. For accurate human perception of audio, audio samples must be presented at a smooth and continuous rate. However, human visual perception is highly tolerant of video quality and frame rate variation; in fact, motion can be per-

ceived even despite a wide variation in video quality and frame rate. Empirical evidence shows that humans can perceive motion if the presentation rate is between 15 and 30 frames/sec. At lower frame rates motion is still perceivable, but artifacts of previous motions are noticeable.

The DVMS of the invention exploits this phenomenon to optimally utilize available computing, compression and network resources; by prioritizing the retrieval, transmission, decompression and presentation of audio over video within a computer system or network computing environment, and by relying on audio-to-video synchronization before presentation, rather than at storage, an acceptable audio rate can be maintained while at the same time varying the video rate to accommodate resource availability in the system or network. Additionally, independent management of audio and video data streams provides many editing capabilities, e.g., the ability to dynamically dub a video stream with multiple audio language streams. Similarly, the synchronized presentation of an audio stream with still pictures is provided for by the independent stream management technique. It must be remembered that all of the synchronization schemes described are applicable to any type of stream, not just audio and video streams.

As described above with reference to Fig. 6, the synchronization of streams within a stream group is the responsibility of the stream interpreter module during a scaling process. The streams may be self-synchronized using either an implicit timing scheme or an explicit timing scheme. Implicit timing is based on the fixed periodicity of the presentation units in the constituent streams of a stream group to be synchronized. In this scheme, each presentation unit is assumed to be of a fixed duration and the presentation time corresponding to each presentation unit is derived relative to a reference presentation starting time. This reference starting time must be common to all of the constituent streams. Explicit timing is based on embedding of presentation time stamps and optionally, presentation duration stamps, within each of the constituent streams themselves and retrieving the stamps during translation of streams from the storage format to the token format. The embedded time stamps are then used explicitly for synchronization of the streams relative to a chosen reference time base.

Using either the implicit or explicit timing self-synchronization schemes, a reference time base is obtained from a reference clock, which advances at a rate termed the reference clock rate. This rate is determined by the reference clock period, which is the granularity of the reference clock ticks.

The DVMS of the invention supports two levels of self-synchronization control, namely, a base level and a flow control level. Base level synchronization is applicable to stream process scenarios in which the stream I/O manager is able to continuously feed stream data to the stream interpreter, without interruption, and in which each presentation unit is available before it is to be con-

sumed. In this scenario, then, the stream I/O manager maintains a process rate and a process work load that guarantees that the stream I/O manager stays ahead of the stream interpreter.

The flow control level of synchronization is a modification of the base level scheme that provides a recovery mechanism from instantaneous occurrences of computational and I/O resource fluctuations which may result in the stream pipe between the stream I/O manager and the stream interpreter running dry. This could occur, for example, in a time-shared or multi-tasked computer environment, in which the stream I/O manager may occasionally fall behind the stream interpreter's demand for presentation units due to a contention, such as a resource or processor contention, with other tasks or with the stream interpreter itself. In such a scenario, the DVMS of the invention augments the base level of synchronization with a stream flow control function, as described below.

Base Level Implicit Timing Synchronization

As explained above, the base level synchronization scheme assumes that there is no need for control of stream flow to the stream interpreter, and thus does not monitor for vacancy of the stream pipe. Implicit timing is based on a reference time base that is applied to each stream to be synchronized.

Considering a scenario in which audio and video streams are to be synchronized, each presentation unit for the video stream to be presented might typically contain video information to be presented in a frame time of, e.g., 33 msec, for NTSC video play. The audio stream might typically be divided into fixed frames of presentation time with marginally varying samples per presentation unit. In a storage scheme in which the audio and video are interleaved, these fixed units of time are set as the time duration for a video frame, i.e., 33 msec.

In this synchronization scenario, the stream interpreter maintains a separate presentation unit counter for each stream pipe, and correspondingly, for each stream in the stream group. The interpreter consumes presentation units from the two streams in a round robin fashion, i.e., first one, then the other, and so on. Importantly, an independent presentation synchronization decision is made for each presentation unit, or token, of each stream, based on a corresponding reference time base, without regard to other streams. This reference time base indicates the current real time relative to the start time of the presentation unit consumption process for the corresponding stream. The stream counter of each stream pipe indicates the number of already consumed presentation units in the corresponding stream. Multiplying this count by the (fixed) duration of each of the presentation units specifies the real time which has elapsed to present the counted units. When this real time product matches the current reference time, the next presentation unit is released for presentation.

The stream interpreter initiates the consumption and presentation of each presentation unit in sequence during its presentation process cycle based on a presentation decision scheme. This scheme implicitly assumes that the stream interpreter is scheduled such that the interpreter process rate is very close to the nominal presentation rate of the corresponding stream. This scheme is based on a comparison of a reference time base with the amount of time required to present the number of already-consumed presentation units, and thus requires the use of counters to keep a count of presentation units as they are consumed.

Base Level Explicit Timing Synchronization

As explained previously, in the explicit timing scheme, stream synchronization is based on time stamps that are embedded in the corresponding streams' tokens themselves. The time stamps represent the time, relative to the reference time base, at which the corresponding audio or video presentation frames are to be consumed and presented. The time base may be, for example, an external clock, or may be generated from the embedded time base of one of the streams to be synchronized. The periodicity of the time stamps is itself flexible and can be varied depending on particular synchronization requirements. Time stamps may be embedded in the streams during capture and compression operations, as described above, or at a later time during, for example, an editing process. Independent of the process by which the time stamps are embedded in a stream, the stamps are utilized by the stream I/O manager and interpreter during playback processes to make the consumption and presentation decisions. The stream interpreter does not maintain a presentation unit counter in this scheme, as it does in the implicit timing scheme. Rather, the embedded time stamps in the streams provide equivalent information.

A time stamp for a presentation frame token consists of two 32-bit integers representing the presentation time and the presentation duration for that presentation unit. The presentation time and the presentation duration are represented in milliseconds. The presentation duration may be omitted if all presentation units are of the same duration.

In this synchronization scheme, the interpreter reads the embedded time stamp of each presentation token, as that token is processed, to determine presentation time and duration for each presentation unit in the sequence. The interpreter decides on consumption and presentation of each presentation unit in each stream based on a decision scheme. This decision scheme is based on the assumption that the stream interpreter is scheduled such that its process rate is very close to the nominal presentation rate of the corresponding stream. This scheme is based on a comparison of a reference time base with the presentation time and presentation duration stamp embedded in each presentation unit.

When a presentation unit's stamp presentation time corresponds to the reference time, that presentation unit is consumed for presentation.

In addition to determining the appropriate time for releasing presentation units in the sequence, both the implicit and explicit timing schemes delete presentation units if the appropriate release time for those units has passed. For example, in the implicit timing scheme, when the product of processed units and unit duration exceeds the currently maintained time count, the next sequential unit is deleted, rather than presented. Similarly, in the explicit timing scheme, then the current presentation time exceeds the time stamp presentation time of a presentation unit, that unit is deleted, rather than presented. In this way, synchronization of streams is maintained, even if units arrive for presentation at a later time than expected.

Flow Control Level Implicit Timing Synchronization

The flow control synchronization scheme augments the base level synchronization scheme to provide for recovery from instantaneous computational and I/O resource fluctuations during a consume and presentation process cycle. The base level scheme relied on the assumption that the stream I/O manager stays ahead of the stream interpreter to keep stream pipes from becoming vacant, or running dry. Flow control synchronization guards against this condition using a scheme based on virtual presentation units.

A virtual presentation unit is one which allows the underlying digital hardware subsystem to continue with a default presentation for the duration of a corresponding presentation unit, while at the same time maintaining a consistent internal state, to thereby provide sequential processing of a stream that is being presented, even while the stream pipe is temporarily empty. Virtual presentation units may be implemented in a variety of embodiments. For example, in the case of motion JPEG video, the playing of a virtual presentation unit would preferably correspond to redisplaying the most recent previous video frame. In the case of audio streams, a virtual presentation unit would preferably correspond to a null unit, i.e., a presentation unit consisting of null samples that represent silence. Other virtual presentation unit implementations are equally applicable.

During a presentation process cycle using the flow control implicit timing scheme to synchronize stream flow, the stream I/O manager and stream interpreter perform the same operations described above in the base level scheme. As explained, the interpreter maintains a separate presentation unit counter for each stream within the stream group being presented, to keep track of the number of already-consumed presentation units in each stream. Multiplying this count by the duration of each presentation unit specifies the time at which, when matching the reference time, the next presentation unit in the sequence is to be presented. The stream inter-

preter decides on the consumption and presentation of each presentation unit based on a decision scheme which assumes that the interpreter is scheduled at a process rate that is close to the nominal stream presentation rate. In this scheme, when the interpreter finds that a presentation token is not available from the stream pipe, and that the reference time and presentation unit count indicate that a presentation unit is needed, a virtual presentation unit is generated and consumed for presentation.

Flow Control Level Explicit Timing Synchronization

During a presentation process cycle using the explicit timing synchronization mechanism augmented with flow control capability, each presentation token in the stream group being presented is assumed to include its own embedded time stamp for presentation time and duration. As in the explicit timing scheme without flow control, the stream interpreter examines each embedded time stamp to decide on the consumption policy of the corresponding presentation unit in the stream pipes set up by the stream I/O manager. The consumption policy is determined based on a decision scheme, which assumes, as did the other schemes, that the process rate of the stream interpreter is close to the nominal presentation rate of the corresponding stream. In this scheme, when it is determined that another presentation unit is not available from the stream pipe and a unit should be presented, a virtual presentation unit is generated based on a default presentation duration, and that unit is then consumed for presentation.

Additionally, in the flow control schemes of either implicit or explicit timing, capability is provided to skip over presentation units. This capability is invoked whenever a previously unavailable presentation unit later becomes available. In the explicit timing scheme, the time stamp of a later available unit will never match the reference time after the presentation of a virtual presentation unit, and thus that unit will never be presented, and will be discarded. In the implicit timing scheme, the presentation of a virtual presentation unit in place of an unavailable presentation unit advances the presentation unit counter, as does any presented unit. When the unavailable unit is then later available, the presentation unit count will be advanced such that the product of the count and the fixed presentation unit duration will not permit presentation of that unit.

Coding of the four synchronization processes described above into instructions suitable for implementing the synchronization techniques will be understandable to those having ordinary skill in the art of C programming.

Self-Synchronization Features

The four self-synchronization schemes described above provide several critical advantages in the digital

vide management scheme of the invention. Self-synchronization accommodates the ability to dynamically associate distinctly stored streams with a common stream group. Thus, for example, audio and video streams may be stored in separate file containers and grouped dynamically during retrieval from storage for synchronized presentation. As discussed above, this synchronization of constituent audio and video streams provides, for example, for the function of dubbing of video with audio, and synchronizing still video with audio. Additionally, using the stream synchronization technique, stream segments from different file containers can be dynamically concatenated into one stream. In the case of explicit self-synchronization, the stream I/O manager marks the first presentation unit in a stream segment with a marker indicating the start of a new stream segment. Then when the stream interpreter consumes this presentation unit, the interpreter reinitializes the reference time base for the corresponding stream.

Self-synchronization further accommodates the ability to adapt to skews in the clock rates of audio and video hardware used to play audio and video streams which are being synchronized. For example, an audio stream recorded at an 11, 22 or 33 KHz sampling rate must be played back at exactly the sampling rate for accurate audio reproduction. Similarly, a video stream recorded at 30 frames per second must be played back at that same rate. The audio and video hardware playing these streams thus must each use clocks adapted for the particular play rate requirement of the corresponding stream. Any skew in the clock rates would cause drifting of the playing streams, and thus destroy synchronization of the streams, if the skew were to be uncorrected. Self-synchronization achieves this correction automatically using a reference time base which the audio and video time bases are checked against; the consumption rate of a stream is adjusted to drop presentation units periodically, if necessary, if a skew in one of the time bases, relative to its prescribed correspondence with the reference time base, is detected, thereby maintaining synchronization with respect to the reference time base and the other stream.

The self-synchronization schemes provide the capability to vary the inherent presentation rate of streams. For example, a video stream captured in PAL format, based on 25 frames per second, may be played in the NTSC format, which is 30 frames per second, albeit with some loss of fidelity. In general, any stream may be played at a custom rate, independent of the rate at which the stream was captured. In fact, it is often desirable in video playback to either speed up or slow down the nominal presentation rate of the video. Using the self-synchronization technique, the video presentation rate may be, for example, sped up by a factor of 2 by simply advancing the reference time base to twice the real time rate. Conversely, the presentation may be slowed by half by advancing the reference time base at one half the real time rate. In these cases, the total time elapsed

for the presentation will be, of course, one half or twice the elapsed time for the presentation made at the nominal rate.

5 Stream Scalability

A scalable stream is a stream that can be played at an aggregate nominal presentation rate with variable data rates, under computer control. Of course, variation in the data rate may affect the quality, fidelity or presentation rate of the stream. The coupling of stream scalability with stream self-synchronization provides a powerful control mechanism for flexible presentation of audio and video stream groups.

As discussed above, scalability allows the DVMS to optimize utility of computer system resources by adjusting stream rates according to utility availability. In the case of audio and video streams, the stream interpreter may be programmed to give higher priority to audio streams than video streams, and thus consume audio presentation units at the nominal audio presentation rate, but consume video units at an available presentation rate. This available presentation rate is determined by the available computational resources of a given computer system. Different computer systems having varying performance characteristics require differing amounts of time to accomplish presentation operations. Such operations involve decompression, format conversion and output device mapping. In particular, a compressed Motion JPEG video stream has to be Huffman decoded, DCT decompressed, converted to RGB color space, and mapped to a 256 color VGA palette by the digital hardware subsystem before presentation within an IBM PC-compatible personal computer system; different computer systems require various time periods to accomplish these tasks. Thus, the management system of the invention adapts to any computer performance characteristics by adjusting the scale of the stream flow rate to accommodate the availability of utilities in that computer.

Most importantly, the stream scalability feature of the digital video management system of the invention provides the ability to comprehensively manage distribution of digital streams over packet networks. The DVMS exploits this capability in a network embodiment providing management protocol schemes for client-server sessions, as well as management protocol schemes for storing, accessing, retrieving and presenting streams over a LAN or WAN. The system thereby accommodates on-demand retrieval and playback of stored streams, and injection and tapping of multicast live streams over packet networks. The managed digital streams may be stored in ordinary computer files on file servers, or may be generated from live analog sources and made accessible over a LAN or WAN. Such access may be on-demand, as mentioned above, as in retrieval and presentation from a stored file, or on-schedule, as in injection and tapping from a broadcast channel. The

management protocol schemes provided by the DVMS will be fully described below.

Referring now to Fig. 9, in a network implementation, the local DVMS manager 20 accesses digital media streams located elsewhere in the network via the remote DVMS manager 82 of the management system; the local DVMS manager provides a client operating environment, while the remote DVMS manager provides a network operating environment. Via the network 80, the local DVMS manager 20 and the remote DVMS manager 82 transmit control messages and digital media data streams as they are requested by a computer client connected in the network.

Remote DVMS Manager

The remote DVMS manager 82 manages network control of digital media streams via four independent modules, namely: a remote stream controller 84, a remote stream input/output (I/O) manager 86, a remote network stream I/O manager 88, and a local network stream I/O manager 90.

In this DVMS network implementation, the local DVMS manager 20, residing locally to a client computer in the network, comprises a local stream controller 24, local stream I/O manager 26 and local stream interpreter 28. The local network stream I/O manager 90 of the remote DVMS manager directly interfaces with the local DVMS manager locally.

The remote stream controller 84 resides on a remote storage device or access point, e.g., a video server, in the network. This controller is responsible for managing the remotely stored streams, e.g., video files, and thereby making them available for on-demand access by the local stream controller module of the local DVMS manager. Client-server session management protocols control this access. The remote stream controller also provides a link for feedback control from the local DVMS manager to the remote DVMS manager, as described below.

The remote stream I/O manager 86 also resides on a remote server; it is responsible for dynamically retrieving and storing streams from or to a storage container in the remote storage server. Efficient access to stored stream information and handling of file formats is provided by this module. Thus, the remote stream I/O manager performs the same tasks as those performed by the stream I/O manager of the local DVMS manager in a stand-alone computer implementation -- tasks including translation between stored stream representations and corresponding dynamic computer-based token representations.

The remote network stream I/O manager 88, implemented on a remote server, regulates transmission of streams across the network to and from a local DVMS manager with which a communications session has been initiated. This transmission comprises stream exchange between the remote network stream I/O man-

ager 88 and the local network stream I/O manager 90, which resides locally with respect to the local DVMS manager modules, on a client in the network. Stream transport protocols control the transmissions. The local network stream I/O manager 90 receives streams from the network and delivers them to the local DVMS stream interpreter 28 during playback processes; conversely, it receives streams from the local stream interpreter and transmits them over the network during recording and storage processes.

The DVMS of the invention provides protocols for managing the interaction and initialization of the local DVMS manager modules and the remote DVMS manager modules just described. Specifically, four classes of protocols are provided, namely, access protocols, for stream group naming and access from a stream server or injector; transport protocols, providing for stream read-ahead, and separation and prioritization of streams; injection/tap protocols, providing the capability to broadcast scheduled streams, e.g., video streams, to selected network clients; and feedback protocols, accommodating the management of adaptive computational resources and communication bandwidths.

When the DVMS is configured in a network environment, remote media data stream file servers in the network advertise the stream groups controlled in their domain based on a standard network advertisement protocol. For example, in the Novell® Netware™ environment, servers advertise based on the Service Advertisement Protocol (SAP). Each video server is responsible for a name space of stream group containers that it advertises.

As shown in Fig. 9, when an application running on a computer (client) connected in the network opens a stream group container by name to access the container contents, the DVMS initializes the corresponding local stream controller 24 of the local DVMS manager to access the corresponding stream group. The local stream controller then sets up a client-server session with the appropriate remote stream controller 82 based on the stream group container name that the application wishes to access and the remote server's advertisement. The local stream controller may access multiple stream group containers during a single session. This capability results from the name service architecture employed by the remote DVMS manager. In this scheme, a domain of container names is accessed via a single access call, whereby multiple containers in the domain are simultaneously available for access.

The local stream controller 24 then initializes the local network stream I/O manager 90 of the remote DVMS manager, and commences a stream read-ahead operation, described below, with the appropriate remote stream controller 84. In turn, that remote stream controller initializes the corresponding remote stream I/O manager 86 and remote network stream I/O manager 88 to handle retrieval and transmission of the constituent streams within the accessed stream group.

The stream read ahead operation is employed to reduce latency perceived by a client when a stream group presentation is begun; stream retrieval, transmission, and scaling require a finite amount of time and would be perceived by a client as a delay. In the read ahead operation, the remote stream I/O manager, the remote network stream I/O manager, and the local network stream I/O manager retrieve, transmit, and scale the streams at the very start of a client-server session, even before the client requests stream presentation. In this scheme, the streams are ready for immediate consumption by the local stream interpreter, via the stream pipes, whenever a user specifies the start of presentation, and possible presentation delays are thereby eliminated or minimized.

Referring now to Fig. 10, when a network client requests access to a specified stream group, the following procedure is implemented. Upon initialization from the request, and based on the network servers' stream group advertisements, the appropriate remote stream I/O manager 86 retrieves stored streams, e.g., audio and video streams, from the appropriate file storage 30 containing the requested stream group. The manager then separates the retrieved streams, if necessary, thereby producing separate audio and video presentation unit streams, and enqueues corresponding stream descriptor tokens in separate stream pipes 87, one pipe for each presentation unit token stream.

The remote network stream I/O manager 88 consumes the presentation unit tokens from each of the stream pipes, assembles transmission packets based on the streams, and releases them for transmission across the network 80 directly to the corresponding local network stream I/O manager 90, based on the DVMS stream data transport protocols; the particular transport protocol used is set by the network environment. For example, in a Novell® network, the Netware SPX protocol is used for stream data transport. The local network stream I/O manager 90, upon receipt of the transmitted presentation units, queues the presentation units in separate stream pipes 32 for each stream to be consumed by the local stream interpreter 28 for use by the client computer's digital media hardware subsystem 34.

Referring to Fig. 11A, illustrating the remote DVMS functions in more detail, upon initialization, the remote stream controller 84 initializes the remote stream I/O manager 86 and the remote network stream I/O manager 88 by creating 130, 136 active modules of each of the managers. It also specifies 132 the requested stream group for access by the two managers. Control 134 of the specified stream group is provided throughout the duration of the managers' functions.

The remote stream controller 84 also provides management 138 of the client/server session which proceeds between the local and remote DVMS systems as a result of the stream group request. Based on information provided by the local DVMS manager which requested the stream group, the remote stream controller

receives 140 a desired rate value from the local DVMS; this rate value indicates the rate at which the streams are to be presented, and is explained more fully below. The remote stream controller specifies 142 this rate to each of the remote stream I/O manager 86 and the remote network stream I/O manager 88, which each receive 144 the rate.

The remote stream I/O manager 86 retrieves, separates, and scales 146 audio and video streams from the appropriate stream container 30. If the streams were stored separately, rather than interleaved, the streams may be individually scaled at this point, while if the streams were interleaved, the remote network stream I/O manager 88 later scales the streams, as explained in detail below.

In a process explained previously with reference to Fig. 7, the remote stream I/O manager creates 148 stream tokens corresponding to the stream presentation unit frames retrieved from storage, and enqueues 150 the stream tokens for delivery to the remote network stream I/O manager via individual stream pipes 32.

The remote network stream I/O manager 88 dequeues 152 the tokens from the stream pipes and if necessary, scales 154 the tokens. The tokens are then formatted 156 for transmission packets, and released to the network for transmission.

Referring also to Fig. 12, the packet format process 156 is implemented as follows. Each token 114 in the token streams 112 is enqueued in a buffer 118, whereby each buffer contains tokens and associated media frame data from one stream only, even if the streams were originally interleaved in storage. Tokens, along with corresponding media data from the buffers, are then sequentially ordered in packets 120 in such a manner that each token and the corresponding media data remain associated. This association, along with the fact that tokens are likely to be time stamped, does not require that the storage format and congruency of the stream be preserved in the transmission packets during transmission.

This packet format scheme provides dramatic advantages over the conventional packet format scheme of the prior art. In the conventional packet protocol the stored media data format, which is typically interleaved, is preserved in the transmission packet format. Thus, in this scheme, audio and video streams are transmitted across a network in packets containing a sequence of interleaved headers, audio frames, and video frames, and thus, the specific syntax by which the interleaved streams were stored is replicated in the packet format.

In contrast, in the packet format scheme of the invention, the separation of streams and distinctly formatting of packets for each stream provides an opportunity and the facility to examine, process, and make transmission decisions about each stream and each presentation unit independent of other streams and presentation units. As a result, the local DVMS manager can make presentation decisions about a given presentation

unit token independent of the other tokens in the corresponding stream, and can make those decisions "on-the-fly". This capability provides for real time scaling and network load adjustment as a stream is retrieved, processed, and transmitted across the network. The conventional prior art scheme does not have any analogous facility, and thus cannot provide the synchronization, scaling, and rate control features of the invention.

Referring to Fig. 11B, once the stream group is transmitted across the network, the local DVMS manager processes the stream group for presentation. The local stream controller 24 manages the client/server session communication with the remote stream controller 84. Like the remote stream controller, it also creates 160, 162 instances of active processors, here initializing the local network stream I/O manager 90 and the local stream interpreter 28. The local stream controller creates 164 the stream grouping of interest and controls 166 that group as the local network stream I/O manager 90 and stream interpreter 28 process the group.

The local network stream I/O manager 90 receives 168 the transmitted network packets and assembles presentation units as they are received. Then it creates 170 stream tokens from the received packets and enqueues 172 them to individual stream pipes. The stream interpreter 28 dequeues 176 the tokens from the stream pipes and scales 176 the tokens as required, in a process discussed below. Then using the synchronization schemes explained previously, the streams are synchronized 178 and sent to the digital hardware subsystem for presentation. The functions of this hardware were explained previously with reference to Fig. 8.

In the reverse process, i.e., when recording streams from a network client for storage on a remote stream server, as shown in Figs. 11A and 11B, the digital stream hardware subsystem provides to the local stream interpreter 28 the stream data, and based on the playing format of the streams, the local stream interpreter generates 180 corresponding time stamps, for use in synchronization and scaling. Stream tokens are then created 182 and enqueued 184 in the stream pipes.

The local network stream I/O manager dequeues 186 the stream tokens from the pipes and scales 188 the streams based on their play rate, record rate, and storage format, as discussed below. Then packets are formed and transmitted 190 via the network to the remote server location on which the corresponding remote DVMS exists.

Thereafter, the remote network stream I/O manager 88 receives 192 the transmitted packets and creates 194 stream tokens based on the packets. The tokens are then enqueued 196 in stream pipes for consumption by the remote stream I/O manager. The remote stream I/O manager dequeues 198 the tokens from the stream pipes, and scales 200 the streams if necessary. Finally, it interleaves the streams, if they are to be stored in an interleaved format, and stores 202 the streams in appropriate stream containers on the server.

Figures 11A and 11B illustrate that the network implementation of the DVMS of the invention is an elegant and efficient extension of the stand-alone DVMS implementation; this extension is possible as a result of the modularity in design of each processing entity. Specifically, the details of packet transport are transparent to the remote stream I/O manager; it functions in the same manner as a stand-alone stream I/O manager. Similarly, presentation unit token streams provided to the local stream interpreter do not contain transmission-specific formats.

As a result, the local DVMS manager, when implemented in a network environment, is easily reconfigured to provide a remote DVMS manager which includes a corresponding remote stream I/O manager, with the addition of a remote network stream I/O manager; and a local DVMS manager which includes a corresponding local stream interpreter, and a local network stream I/O manager from the remote DVMS manager. Exploiting this modularity, programming applications may be created which are supported by the DVMS functionality without them perceiving a functional difference between a local, stand-alone type stream scenario and a remote, network stream scenario.

Additionally, as will be recognized by those skilled in the art, these processes may alternatively be implemented in hardware using standard design techniques to provide the identical functionality.

30 Scalable Stream Rate Control

In the network embodiment of the DVMS of the invention, the remote and local DVMS managers operate together to provide control of the rate of flow of streams through a network during stream transmission. As mentioned above, this capability is particularly advantageous in handling audio and video streams to accommodate fluctuations in network utility availability by prioritizing audio stream rate over video stream rate.

This priority is based on the premise that human visual perception of motion is highly tolerant of variations in the displayed quality and frame rate of presented video. Typically, humans perceive motion when a video presentation rate exceeds at least 15 frames per second. Moreover, instantaneous and smooth variations in video presentation rates are practically unnoticeable. However, human aural perception is quite intolerant of variations in audio presentation quality or rate. Typically, humans perceive noise when a constant audio presentation rate is not maintained, and perceive "clicks" when brief periods of silence are injected into an audio stream. Thus, the DVMS system prioritizes audio streams over video streams. This prioritization of audio over video extends over the entire data flow of audio and video streams in a network, starting from their retrieval from storage containers and ending with their presentation.

Control of the rate of streams through a network based on this audio prioritization scheme may be initi-

ated automatically, or in response to a direct user request. Each type of control request is discussed below in turn. The remote DVMS manager responds to each type in the same manner, however.

Referring again to Fig. 11A, remote stream controllers 84 in the network are responsible for instructing the corresponding remote stream I/O manager 86 and remote network stream I/O manager 88 as to what percentage of the nominal presentation rate (at which the stream would "normally" be presented) the stream should be actually retrieved and transmitted. The remote stream controller receives 140 the desired rate value via network communication with the local stream controller 24 and specifies 142 this rate to the remote stream I/O manager 86 and the remote network stream I/O manager 88, which each receive 144 the rate value.

The stream rate control mechanism is carried out by either the remote stream I/O manager or the remote network stream I/O manager, depending on particular stream access scenarios. As explained above, if the requested audio and video streams are interleaved in storage, in, e.g., the Intel DVI AVSS file format, the remote stream I/O manager retrieves the streams in that interleaved form, separates the streams into distinct streams, and creates corresponding presentation unit tokens. The remote stream I/O manager does not, in this scenario, have the ability to manipulate the streams distinctly because they are retrieved interleaved. In this case, the remote network stream I/O manager, which obtains the streams from the stream pipe after they have been separated, controls the rate of each stream as before forming stream packets for network transmission.

If the streams to be retrieved are individually stored, the remote stream I/O manager may control the rate of the streams as they are each separately retrieved and corresponding tokens are created. In this case, the rate control functionality of the remote network stream I/O manager is redundant and does not further change the stream rate before the stream is transmitted across the network.

Rate control of noninterleaved streams is provided by the remote stream I/O manager during the scaling process 146, in which case the remote stream I/O manager retrieves stream frames from the storage container while skipping over appropriate stream frames to achieve the prespecified stream rate. The streams frames which are skipped over are determined based on the particular compression technology that was applied to the stream. The remote stream I/O manager substitutes virtual presentation units for the skipped stream frames to maintain sequential continuity of the stream.

As explained previously regarding flow control synchronization schemes, a virtual presentation unit comprises a presentation unit with some amount of substitute media data information for maintaining a consistent internal state of stream unit sequence, even while a next sequential unit is unavailable. Here in the case of scal-

ing, where virtual units are employed to scale the transmission rate of streams, virtual units are additionally employed to reduce the amount of presentation unit data that is transmitted.

Accordingly, here a virtual video presentation unit comprises a null presentation unit, having a specified presentation duration and time, or a time stamp, but not containing any frame presentation information. Then, when the remote stream I/O manager substitutes a virtual presentation unit for a skipped stream frame, a transmission packet including the virtual presentation unit is shorter and more quickly transmitted than it would be if the skipped frame was included. When the local stream interpreter and digital presentation subsystem receive and process the null video unit, they interpret that unit as an instruction to represent the most recently presented frame. In this way, the presentation subsystem maintains default video presentation data without requiring that data to be received via a network transmission.

As will be recognized by those skilled in the art of compression technology, it is alternatively possible, using appropriate compression techniques, to substitute partial media information, rather than null information to increase or decrease the transmission rate of presentation streams containing presentation units that will not be presented.

Rate control of interleaved streams is provided by the remote network stream I/O manager upon receipt of the stream tokens from the stream pipes. Here, the remote network stream I/O manager scales 154 the stream tokens as they are processed to form transmittal packets. This is accomplished by processing the stream in a scheme whereby the remote network stream I/O manager skips over appropriate tokens and substitutes virtual presentation unit tokens in their place, depending on the compression technology used, to achieve the specified stream rate.

In this common and important situation of interleaved stream storage, the remote network stream I/O manager participates in stream data flow and thus may be characterized with a particular process cycle and process period. During each of its process cycles, the remote network stream I/O manager processes a single presentation unit and determines if the next sequential presentation unit is to be transmitted based on a transmit decision scheme. Like the process decision schemes described above in connection with synchronization techniques, the transmit decision scheme is implemented based on the timing technique of the stream being processed; if the stream presentation units include embedded time stamps, then the transmit decision scheme is based on an explicit timing count, while implicit timing counting is employed otherwise.

No matter which agent provides the scaling function, only video streams are scaled, while audio stream presentation frames and tokens are processed at the full nominal presentation rate, without skipping any au-

audio presentation frames; this preservation of audio presentation rate inherently prioritizes audio streams over video streams.

The scaling function is, as explained above, dependent on the compression technology employed for a particular frame or stream group. Using, e.g., a key frame-based compression technique, a key frame is an independently selectable frame within a stream that contains information required for decompression of all the following non-key frames dependent on that key frame. Dependent, or non-key, frames are not independently selectable. The motion JPEG format relies on a scheme in which every frame in a stream is a key frame. During the scaling operation, only key frames are skipped over, whereby all non-key frames associated with the skipped key frame are also skipped over. Null frames are then substituted for the key frame and all of its corresponding non-key frames.

An implicit timing rate control scheme and an explicit timing rate control scheme may be devised. Like the synchronization techniques described previously, the implicit rate control scheme is based on a counting technique and does not require embedded time codes on the stream presentation frames. The explicit rate control scheme is based on the use of time stamps for explicitly determining the presentation and duration time of a given frame. In either implementation, virtual presentation units are generated to accommodate skipped stream frames.

In addition, in either implementation, when skipped stream frames later become available, they are identified and skipped over, thereby being deleted, rather than presented. This presentation unit deletion function, like that employed in the synchronization schemes, maintains a current sequential stream progression.

Adaptive Load Balancing

The DVMS of the invention includes the ability to automatically and dynamically sense the load of a packet network in which the system is implemented. Based on the sensed loading, the stream rate control mechanism described above is employed by the system to correspondingly and adaptively balance the load within the network, thereby optimizing the network utility availability.

Referring to Fig. 11B, in this load balancing scheme, the local network stream I/O manager 90 monitors 206 the stream pipes 32 currently transmitting streams between that manager and the local stream interpreter 28 for variations in the average queue size, i.e., availability of presentation unit tokens, of each pipe. When the average queue size varies significantly, the local network stream I/O manager detects the direction of the change, i.e., larger or smaller. Thereafter, it notifies 208 the local stream controller 24 of the change and requests a new stream presentation token rate to be transmitted as a percentage of the nominal presentation

rate, based on the change. In turn, the local stream controller transmits the request to the remote stream controller 84, which in response, instructs the remote stream I/O manager 86 and the remote network stream I/O manager 88 to adjust the stream presentation unit rate to the requested rate.

The requested rate is based on the average queue size in the following scheme. When the queue size increases significantly above a prespecified upper availability, the requested rate is increased; the increased availability indicates that high-speed processing may be accommodated. Conversely, when the queue size decreases significantly below a prespecified lower availability, the requested rate is decreased; the decreased availability indicates that the current rate cannot be accommodated and that a lower rate is preferable.

Alternatively, a user may specify a desired stream presentation rate, that specification being accepted by the local stream controller 24. In turn, the local stream controller sends the request to the remote stream controller for implementation.

In the corresponding reverse process, in which stream frames are stored after being recorded via the local DVMS manager, the remote stream I/O manager scales 200 the stream before storage to reconstruct the stream such that it no longer includes null frames. This function may also be accomplished by the local network stream I/O manager in a scaling process 188 completed before a stream is transmitted.

The DVMS of the invention has been described with particular detail relating to a preferred embodiment. Other embodiments are intended to fall within the scope of the invention. For example, while the DVMS of the invention has been described in a scheme for managing audio and video streams, other media data stream types, e.g., stills, accessed from various media data access points, e.g., a PBX server, are within the scope of the claims. If the DVMS is implemented on a computer system or network in software, programming languages other than the C programming language may be employed, as will be clear to those skilled in the art of programming. Alternatively, the DVMS may be implemented entirely in hardware using standard digital design techniques, as will also be clear to those skilled in the art of digital hardware design.

Claims

1. A computer-based media data processor for controlling the timing of computer processing of digitized continuous time-based media data composed of a sequence of presentation units, each unit having a prespecified presentation duration during a computer presentation of the media data, the media data processor comprising:

a reference clock which indicates a start time

of presentation processing of the media data presentation units and which maintains a current presentation time as the media data presentation unit sequence is processed for presentation;

a counter for counting each presentation unit in the presentation unit sequence after that presentation unit is processed for presentation to maintain a current presentation unit count; and a comparator connected to the reference clock and the counter, and programmed with the prespecified presentation duration, the comparator comparing a product of the presentation duration and the current presentation unit count, specified by the counter, with the current presentation time, specified by the reference clock, after each presentation unit is processed for presentation and based on the comparison, releasing a next sequential presentation unit to be processed for presentation when the product matches the current presentation time count, and deleting a next sequential presentation descriptor in that sequence when the product exceeds the current presentation time count.

2. The media data processor of claim 1 further comprising a flow controller, connected to said comparator, for receiving an indication from the comparator that a presentation unit should be released for presentation, determining availability of a next presentation unit in the presentation unit sequence to be processed, and based on that availability, generating and releasing a virtual presentation unit of the prespecified presentation duration to be presented as a default presentation unit in place of a next presentation unit when a next presentation unit is not available and until the next presentation unit is available.

3. The media data processor of claim 2 wherein the flow controller is adapted to monitor and identify a previously unavailable presentation unit when that unit is later available and in response to identification of the later available unit, withholding the unit from release from presentation, whereby the later available unit is not presented.

4. The media data processor of claim 1 wherein:

the clock is adapted to indicate a start time of presentation processing of a plurality of media data presentation unit sequences, the start time being common to the plurality of sequences, and which maintains a current presentation time as the media data sequences are processed for presentation;

the counter counts each presentation unit in the

plurality of presentation unit sequences after that presentation unit is processed for presentation to maintain a distinct current presentation unit count for each presentation unit sequence; and

the comparator is connected to the reference clock and the counter, and programmed with the prespecified presentation duration, the comparator comparing for each of the plurality of presentation unit sequences a product of the presentation unit duration and the current presentation unit count of that sequence, specified by the counter, with the current presentation time, specified by the reference clock, after each presentation unit from that sequence is processed for presentation, and based on the comparison, releasing a next sequential presentation unit in that presentation unit sequence to be processed for presentation when the product matches the current presentation time count, and deleting a next sequential presentation unit in that presentation unit sequence when the product exceeds the current presentation time count, whereby the plurality of media data presentation unit sequences are synchronously processed for substantially simultaneous synchronous presentation.

5. A computer-based media data processor for controlling the computer presentation of digitized continuous time-based media data composed of a sequence of presentation units, each unit having a prespecified presentation duration and presentation time during a computer presentation of the media data and further characterized as a distinct media data type, the media data processor comprising:

a media data input manager for retrieving media data from a corresponding media data access location in response to a request for computer presentation of specified presentation unit sequences, determining the media data type of each presentation unit in the retrieved media data, designating each retrieved presentation unit to a specified media data presentation unit sequence based on the media data type determination for the presentation unit, assembling a sequence of presentation descriptors for each of the specified presentation unit sequences, each presentation descriptor comprising media data for one designated presentation unit in that sequence, all presentation descriptors in an assembled sequence being of a common media data type, and linking the presentation descriptors in each assembled sequence to establish a progression of presentation units in each of the sequences; and a media data interpreter, connected to the me-

media data input manager, for accepting from the media data input manager the assembled presentation descriptor sequences one descriptor at a time and releasing the sequences for presentation one presentation unit at a time, indicating a start time of presentation processing of the presentation unit sequences, maintaining a current presentation time as the sequences are processed for presentation, counting each unit in the sequences after that unit is released to be processed for presentation, to maintain a distinct current presentation unit count for each sequence, comparing for each of the presentation unit sequences a product of the presentation duration and the current presentation unit count of that sequence with the currently maintained presentation time after each unit from that sequence is processed for presentation, and based on the comparison, releasing for presentation processing a next sequential presentation unit in that sequence when the product matches the currently maintained presentation time count, and deleting a next sequential presentation unit in that sequence when the product exceeds the currently maintained presentation time count.

6. The media data processor of claim 5 further comprising a presentation unit sequence controller for initiating the media data input manager and the media data interpreter, specifying to the media data input manager and the media data interpreter the presentation unit sequences to be presented, and controlling starting and stopping of sequence presentation in response to user specification.
7. The media data processor of claim 5 wherein the media data retrieved by the media data input manager comprises a storage presentation unit sequence composed of presentation units for the specified presentation unit sequences, presentation units of the specified presentation unit sequences being alternately interleaved to compose the storage presentation unit sequence.
8. The media data processor of claim 5 wherein the media data retrieved by the media data input manager comprises a plurality of storage presentation unit sequences, each storage presentation unit sequence composed of presentation units for a specified presentation unit sequence and all presentation units in a storage presentation unit sequence being of a common media data type.
9. The media data processor of claim 6 wherein the retrieved media data presentation units are encoded in a storage code and compressed, and further comprising a presentation system for decoding the

presentation units, decompressing the presentation units, and converting the digitized presentation units to a corresponding analog representation for presentation.

10. The media data processor of claim 5 wherein the media data interpreter maintains the current presentation time at a prespecified time rate such that presentation units of the specified presentation sequences are each presented for a presentation duration different than the prespecified presentation duration.
11. The media data processor of claim 5 wherein the media data interpreter monitors for each specified presentation unit sequence an actual presentation rate of the presentation descriptors in that sequence released for presentation, compares the actual presentation rate with a prespecified nominal presentation rate, and based on the comparison, generates and releases a virtual presentation unit of the prespecified presentation duration to be presented as a default presentation unit each time the monitored presentation rate is greater than the prespecified presentation rate, and based on the comparison, skips over a presentation unit each time the monitored presentation rate is less than the prespecified presentation rate.
12. A computer-based media data processor for controlling transmission of digitized media data in a packet switching network, the media data comprising a sequence of continuous time-based presentation units, each unit having a prespecified presentation duration and presentation time during a computer presentation of the media data and further being a distinct media data type, the network comprising a plurality of client computer processing nodes interconnected via packet-based data distribution channels, the media data processor comprising:
 - a remote media data controller for receiving from a client processing node a request for presentation of specified presentation unit sequences;
 - a remote media data input manager for receiving from the media data controller an indication of the specified presentation unit sequences, and in response to the request, retrieving media data from a corresponding media access location, determining the media data type of each presentation unit in the retrieved media data, designating each retrieved presentation unit to a specified media data presentation unit sequence based on the media data type determination for the presentation unit, assembling a sequence of presentation descriptors for each

of the specified presentation unit sequences, each descriptor comprising media data for one designated presentation unit in that sequence, all presentation descriptors in an assembled sequence being of a common media data type, and linking the descriptors in each assembled sequence to establish a progression of presentation units in each of the specified presentation unit sequences;

a remote network media data manager connected to the remote media data input manager, for accepting from the remote media data input manager the assembled specified presentation descriptor sequences one presentation descriptor at a time, assembling transmission presentation unit packets each composed of at least a portion of a presentation descriptor and its media data, all presentation descriptors and media data in an assembled packet being of a common media data type, and releasing the assembled packets for transmission via the network to the client processing node requesting presentation of the specified presentation unit sequences;

a local media data controller for transmitting the request for presentation of specified presentation unit sequences from the client processing node to the remote media data input controller via the network and controlling starting and stopping of sequence presentation in response to user specifications;

a local network media data input manager for receiving from the local media data controller an indication of the specified presentation unit sequences, receiving the transmission presentation unit packets transmitted from the remote network media data manager via the network, designating a presentation unit sequence for each presentation descriptor and media data in the received packets to thereby assemble the presentation descriptor sequences each corresponding to one specified presentation unit sequence, all presentation descriptors and media data in an assembled sequence being of a common media data type, and linking the descriptors in each assembled sequence to establish a progression of presentation units for each of the presentation unit sequences; and

a local media data interpreter, connected to the local network media data input manager, for accepting the assembled presentation descriptor sequences one descriptor at a time and releasing the sequences for presentation one unit at a time, indicating a start time of presentation processing of the sequences, maintaining a current presentation time as the descriptor sequences are processed for presentation, and based on the presentation duration of each

presentation unit, synchronizing presentation of the specified presentation unit sequences with the current presentation time, wherein the local media data interpreter synchronizes presentation of the specified presentation unit sequences by counting each presentation unit in the sequences after that presentation unit is released to be processed for presentation to maintain a distinct current presentation unit count for each sequence, comparing for each of the presentation unit sequences a product of the presentation duration and the current presentation unit count of that sequence with the maintained presentation time after a presentation unit from that sequence is released to be processed for presentation, and based on the comparison, releasing a next sequential presentation unit in that presentation unit sequence when the product matches the currently maintained presentation time, and deleting a next sequential presentation unit in that presentation unit sequence when the product exceeds the currently maintained presentation time.

13. The media data processor of claim 8 or 12 wherein at least one media data input manager associates each presentation descriptor with a corresponding presentation duration and presentation time, based on the retrieved media data.
14. The media data processor of claim 1, 5 or 12 wherein the media data presentation unit sequence comprises a video frame sequence including a plurality of intracoded video frames.
15. The media data processor of claim 14 wherein each of the plurality of intracoded video frames comprises a key frame and is followed by a plurality of corresponding non-key frames, each key frame including media data information required for presentation of the following corresponding non-key frames.
16. The media data processor of claim 12 wherein the local media data interpreter synchronizes presentation of the specified presentation unit sequences by comparing for each of the presentation descriptors in each of the presentation descriptor sequences the presentation time corresponding to that descriptor with the currently maintained presentation time, and based on the comparison, releasing a next sequential presentation unit to be processed for presentation when the corresponding presentation time of that descriptor matches the current presentation time, and deleting a next sequential presentation unit to be processed for presentation when the current presentation time exceeds the corresponding presentation time of that descriptor.